

EDUTORIAL

Retrospective Studies

A retrospective study investigates outcomes specified at the beginning of a study by looking backwards at data collected from previous patients. Patients are enrolled after the clinical event of interest or exposure has occurred: this is usually conducted by review of the medical notes. Retrospective studies may be either cohort or case–control studies and have four primary purposes: (1) either as an audit tool for comparison of the historical data with current or future practice, (2) to test a potential hypothesis regarding suspected risk factors in relation to an outcome, (3) to ascertain the sample size and data required for a prospective study or trial, or (4) to investigate uncommon or rare events (e.g. graft infection), where the size of a prospective study would be prohibitively large and take too long to conduct: here a case–control design may be efficient. The validity of previous interventions from either patient case notes or hospital records will be reasonably reliable, but other exposures, such as smoking and dietary history or specific drug prescriptions, may be less reliable and suffer from recall bias or are simply not available. The other main disadvantage of retrospective studies is the potential for selection bias in how controls in a case–control study are ascertained.

A few simple steps will enhance the quality of retrospective studies.

- 1 The aims and objectives of the study to test *specific hypotheses* should be specified in advance. Develop and refine your hypothesis with clinical colleagues.
- 2 Conduct a *literature review*, preferably a systematic review, to assess what is already known on the topic.
- 3 Write a *formal research proposal*, with introduction, aims, methods, and sample size estimation. The selection (inclusion and exclusion criteria), size of target population and the choice of which variables to measure (outcomes, exposures, and potential confounders) should be specified and appropriate case record forms (CRFs), paper or electronic, designed, with codes specified for missing data. Check that the relevant variables usually are documented in the medical records. Sample size should be based on the anticipated outcomes (using information coming from the review conducted in 2 above), rather than from the number of cases exposed to the treatment/environment/gene of interest. This research proposal should be discussed with and reviewed by colleagues.
- 4 The reviewed research proposal should be discussed with your *Institutional Ethics or Review* panel for their formal approval.
- 5 *Selection and recruitment of patients* needs careful consideration, to avoid bias. In a case–cohort study, cases who have been exposed to the risk factor of interest may be more likely to enrol in the study (e.g. if they have a vested interest in the results of the study), which can bias results, whereas the controls recruited need to be from the same population as the

cases (often done by matching controls to cases using key demographics such as age and sex and possibly other important information such as workplace restrictions if assessing an occupational exposure).

- 6 *Data* should be abstracted only on to the CRFs. At least a sample of the data should be extracted by two people to ensure consensus. For this you may need to write a coding manual, e.g. for aspirin use code 1 for yes, 2 for no, 3 for discontinued, 4 for unknown or missing. Try a pilot study, with subsequent revision of the CRF and coding manual. Check for interobserver reliability/consensus for data extraction: using data extractors who are blind to the hypothesis being investigated is a helpful tool for reducing bias. It may be possible to undertake these checks, using anonymized records, while waiting for ethical review.
- 7 In *analysis*, investigators need to measure and control for potential confounders using appropriate statistical methodology before reliably interpreting results. Appropriate statistical techniques should be used for matched case–control studies (e.g. conditional logistic regression). It also may be necessary to account for missing data using appropriate methodology (e.g. multiple imputation). Careful consideration before beginning the study must be taken to identify and subsequently measure important prognostic variables that may differ between the exposure groups: differences in these known characteristics need to be adjusted for in the final analysis. Stratification (subgrouping) and adjustment with multivariate regression models are the two most common statistical techniques employed, but neither of these techniques can eliminate bias related to unmeasured or unknown confounders.
- 8 *Reporting* of retrospective cohort and case–control studies should follow STROBE guidelines. For instance, length of follow up should always be reported. Reading these guidelines before starting a study may help improve the study design.

FURTHER READING

- 1 Vassar M, Holzmann M. The retrospective chart review: important methodological considerations. *J Educ Eval Health Prof* 2013;10:12.
- 2 *Strengthening the reporting of observational studies. STROBE statement*. Retrieved July 13 2015 from: <http://www.strobe-statement.org>.

J.T. Powell*

Vascular Surgery Research Group, Imperial College,
London W6 8RP, UK

M.J. Sweeting

Department of Public Health and Primary Care,
University of Cambridge, Cambridge CB1 8RN, UK

*Corresponding author. Vascular Surgery Research Group,
Imperial College, London W6 8RP, UK.
Email-address: j.powell@imperial.ac.uk (J.T. Powell)